



# Business Statistics: A Decision-Making Approach

7<sup>th</sup> Edition

---

## **Chapter 2**

### **Graphs, Charts, and Tables – Describing Your Data**



# Chapter Goals

---

**After completing this chapter, you should be able to:**

- Construct a frequency distribution both manually and with a computer
- Construct and interpret a histogram
- Create and interpret bar charts, pie charts, and stem-and-leaf diagrams
- Present and interpret data in line charts and scatter diagrams



# Frequency Distributions

---

## What is a Frequency Distribution?

- A frequency distribution is a **list or a table** ...
- containing the **values of a variable** (or a set of ranges within which the data fall) ...
- and the **corresponding frequencies** with which each value occurs (or frequencies with which data fall within each range)



# Why Use Frequency Distributions?

---

- A frequency distribution is a way to summarize data
- The distribution condenses the raw data into a more useful form...
- and allows for a quick visual interpretation of the data

# Frequency Distribution: Discrete Data

- **Discrete data:** possible values are countable

**Example:** An advertiser asks 200 customers how many days per week they read the daily newspaper.



Number of days read	Frequency
0	44
1	24
2	18
3	16
4	20
5	22
6	26
7	30
<b>Total</b>	<b>200</b>



# Relative Frequency

**Relative Frequency:** What proportion is in each category?

Number of days read	Frequency	Relative Frequency
0	44	.22
1	24	.12
2	18	.09
3	16	.08
4	20	.10
5	22	.11
6	26	.13
7	30	.15
<b>Total</b>	<b>200</b>	<b>1.00</b>

$$\frac{44}{200} = .22$$

22% of the people in the sample report that they read the newspaper 0 days per week





# Frequency Distribution: Continuous Data

---

- **Continuous Data:** may take on any value in some interval

**Example:** A manufacturer of insulation randomly selects 20 winter days and records the **daily high temperature**

**24, 35, 17, 21, 24, 37, 26, 46, 58, 30,  
32, 13, 12, 38, 41, 43, 44, 27, 53, 27**

(Temperature is a continuous variable because it could be measured to any degree of precision desired)



# Grouping Data by Classes

---

Sort raw data from low to high:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

- Find range:  $58 - 12 = 46$
- Select number of classes: 5 (usually between 5 and 20)
- Compute class width: 10 (46/5 then round off)
- Determine class boundaries: 10, 20, 30, 40, 50  
(Sometimes class midpoints are reported: 15, 25, 35, 45, 55)
- Count the number of values in each class





# Frequency Distribution Example

**Data from low to high:**

**12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58**

<b>Frequency Distribution</b>		
<b>Class</b>	<b>Frequency</b>	<b>Relative Frequency</b>
<b>10 but under 20</b>	<b>3</b>	<b>.15</b>
<b>20 but under 30</b>	<b>6</b>	<b>.30</b>
<b>30 but under 40</b>	<b>5</b>	<b>.25</b>
<b>40 but under 50</b>	<b>4</b>	<b>.20</b>
<b>50 but under 60</b>	<b>2</b>	<b>.10</b>
<b>Total</b>	<b>20</b>	<b>1.00</b>



# Frequency Histograms

---

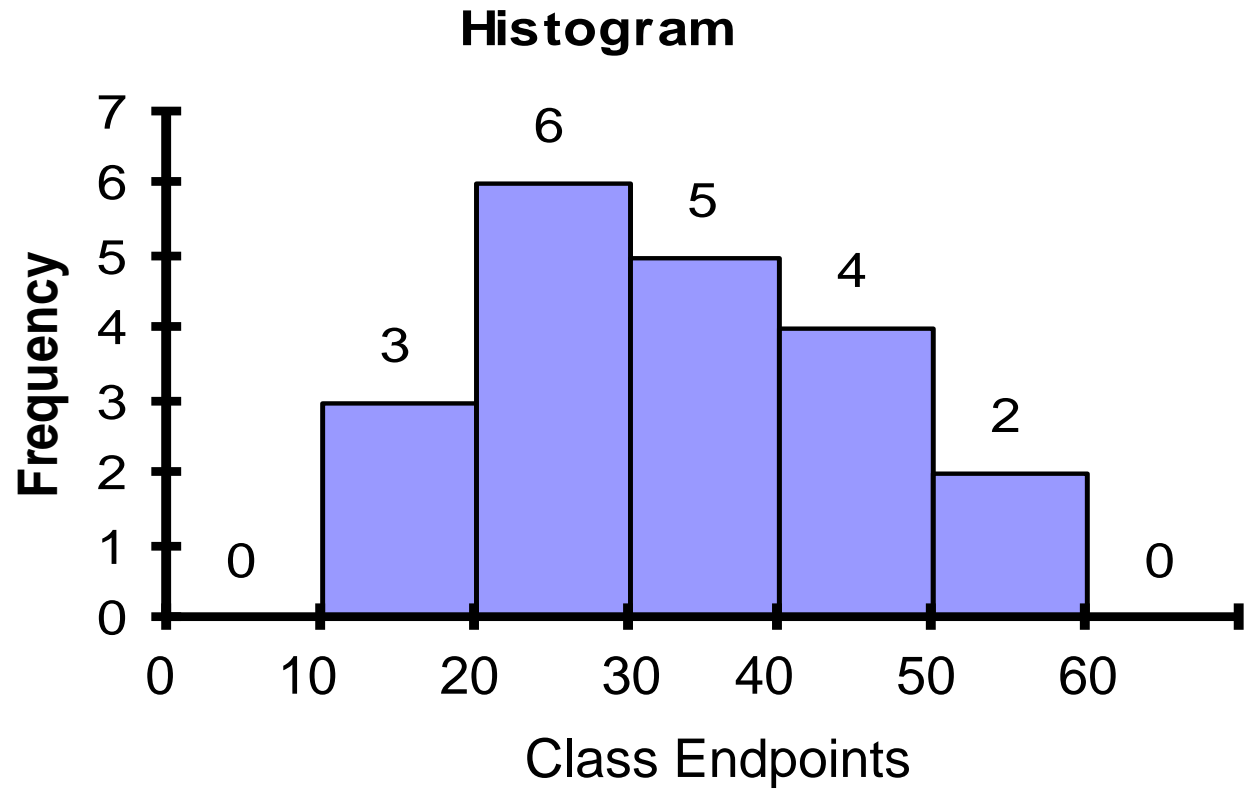
- The **classes** or **intervals** are shown on the horizontal axis
- **frequency** is measured on the vertical axis
- Bars of the appropriate heights can be used to represent the number of observations within each class
- Such a graph is called a **histogram**



# Histogram Example

**Data in ordered array:**

**12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58**



No gaps between bars, since continuous data



# Questions for Grouping Data into Classes

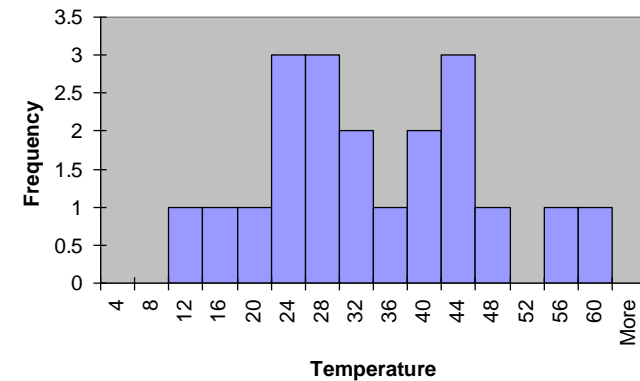
---

- 1. How wide should each interval be?  
(How many classes should be used?)
- 2. How should the endpoints of the intervals be determined?
  - Often answered by trial and error, subject to user judgment
  - The goal is to create a distribution that is neither too "**jagged**" nor too "**blocky**"
  - Goal is to appropriately show the pattern of variation in the data

# How Many Class Intervals?

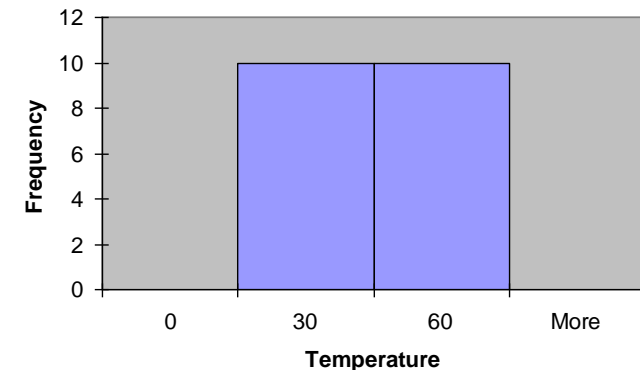
## ■ Many (Narrow class intervals)

- may yield a very jagged distribution with gaps from empty classes
- Can give a poor indication of how frequency varies across classes



## ■ Few (Wide class intervals)

- may compress variation too much and yield a blocky distribution
- can obscure important patterns of variation.



(X axis labels are upper class endpoints)



# General Guidelines

<i>Number of Data Points</i>	<i>Number of Classes</i>
under 50	5 - 7
50 – 100	6 - 10
100 – 250	7 - 12
over 250	10 - 20

- Class widths can typically be reduced as the number of observations increases
- Distributions with numerous observations are more likely to be smooth and have gaps filled since data are plentiful



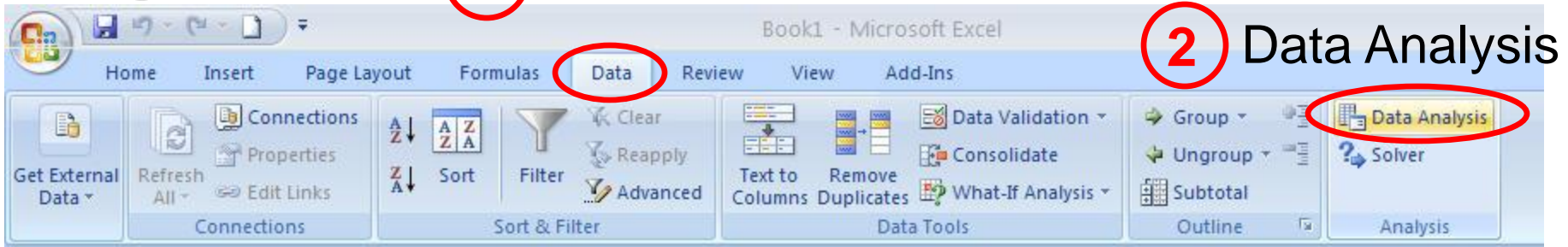
# Class Width

- The class width is the distance between the lowest possible value and the highest possible value for a frequency class
- The class width is

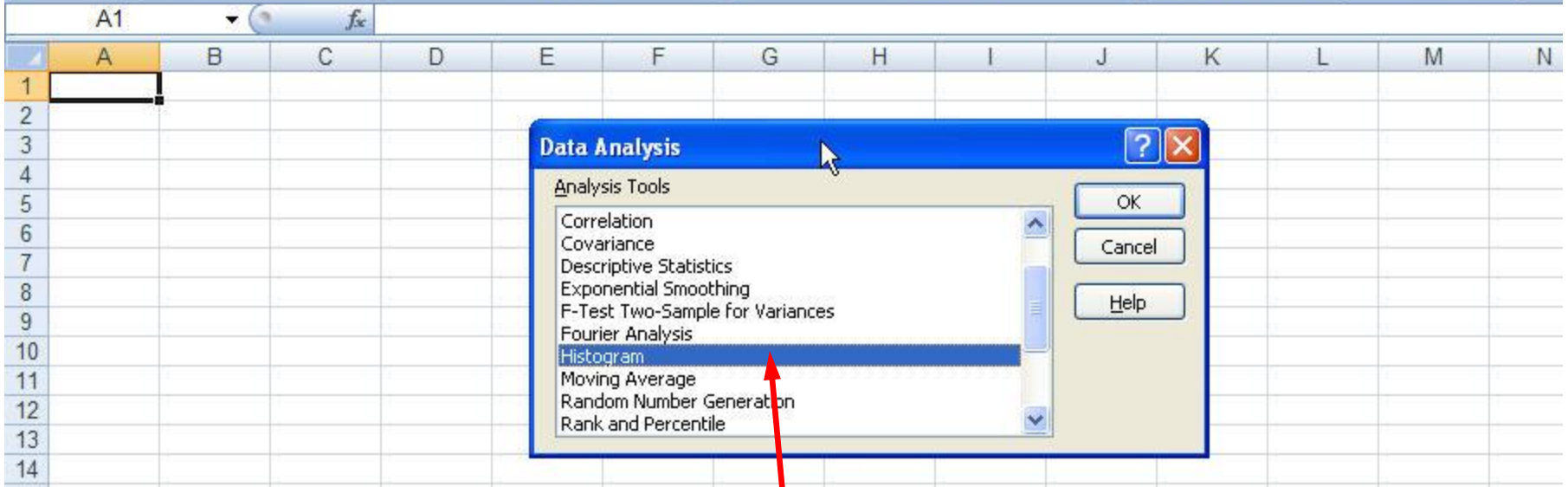
$$W = \frac{\text{Largest Value} - \text{Smallest Value}}{\text{Number of Classes}}$$

# Histograms in Excel

1 Select "Data" Tab



2 Data Analysis



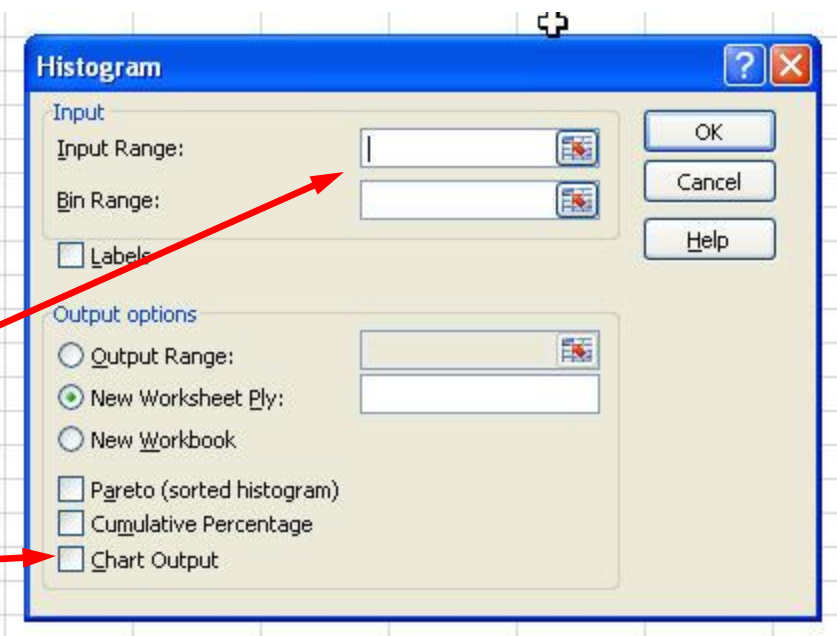
3 Choose Histogram





# Histograms in Excel

*(continued)*



4

Input data and bin ranges

Select Chart Output



# Ogives

---

- An **Ogive** is a graph of the **cumulative relative frequencies** from a relative frequency distribution
- Ogives are sometime shown in the same graph as a relative frequency histogram



# Ogives

(continued)

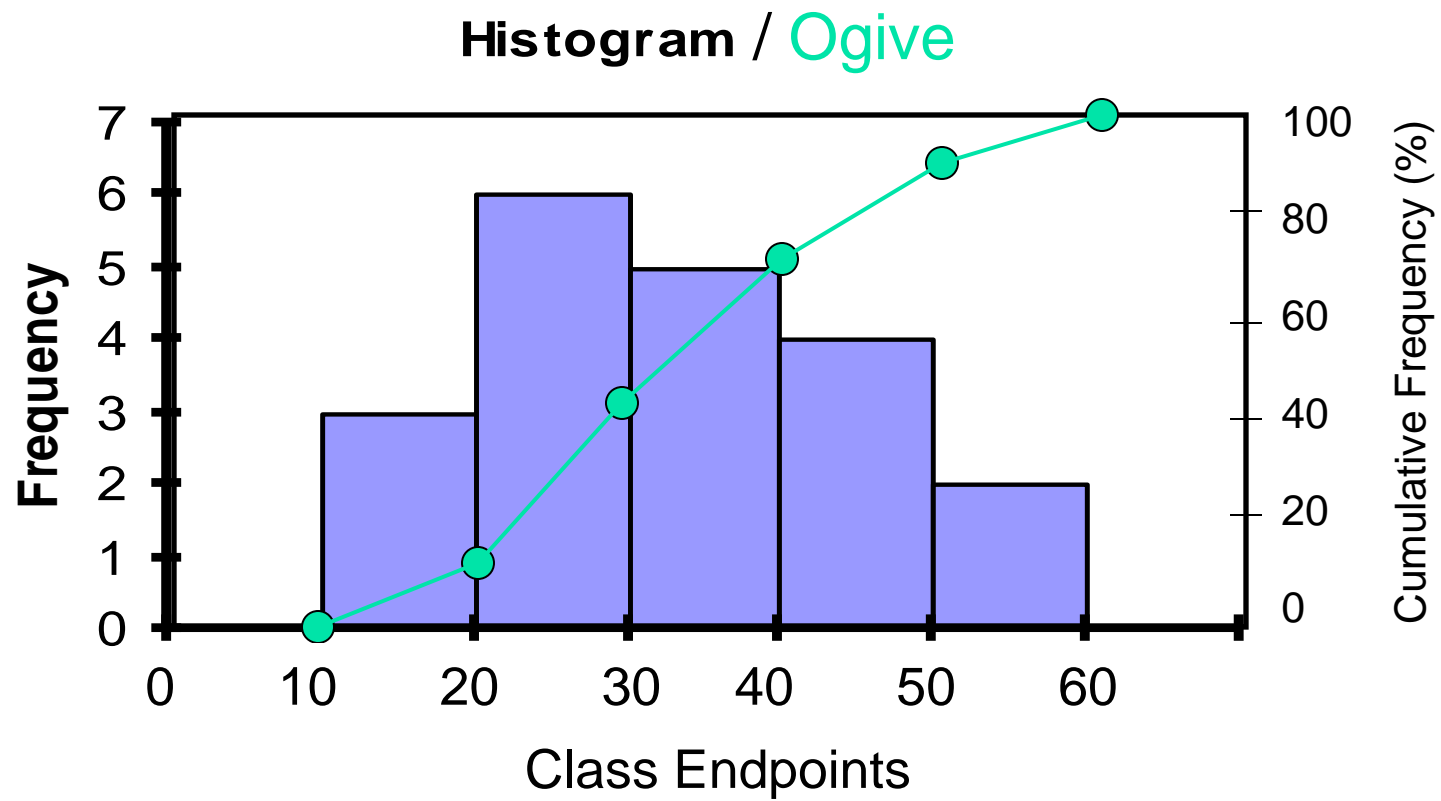
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Add a cumulative relative frequency column:

Frequency Distribution			
Class	Frequency	Relative Frequency	Cumulative Relative Frequency
10 but under 20	3	.15	.15
20 but under 30	6	.30	.45
30 but under 40	5	.25	.70
40 but under 50	4	.20	.90
50 but under 60	2	.10	1.00
<b>Total</b>	<b>20</b>	<b>1.00</b>	



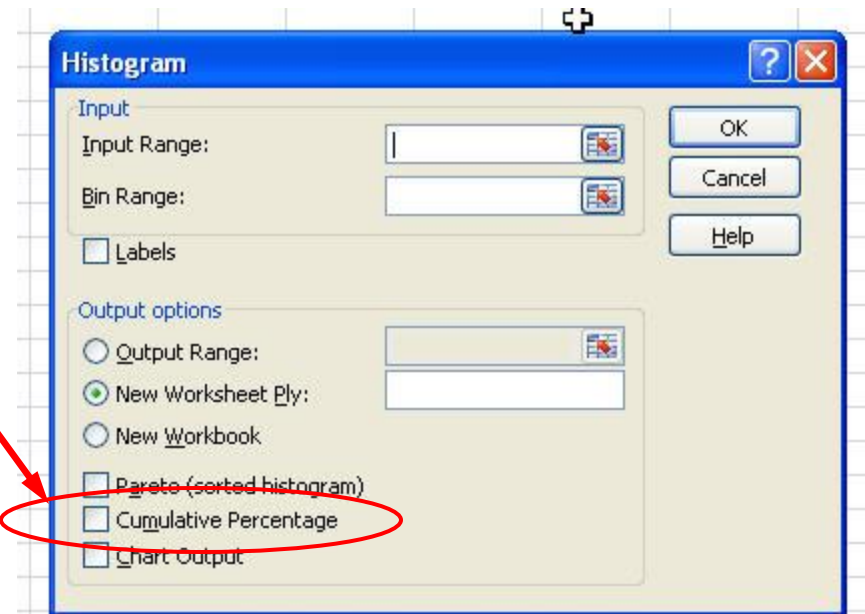
# Ogive Example





# Ogives in Excel

Excel will show the Ogive graphically if the “Cumulative Percentage” option is selected in the Histogram dialog box



# Other Graphical Presentation Tools



**Categorical  
Data**

**Bar  
Chart**

**Pie  
Charts**

**Quantitative  
Data**

**Stem and Leaf  
Diagram**



# Bar and Pie Charts

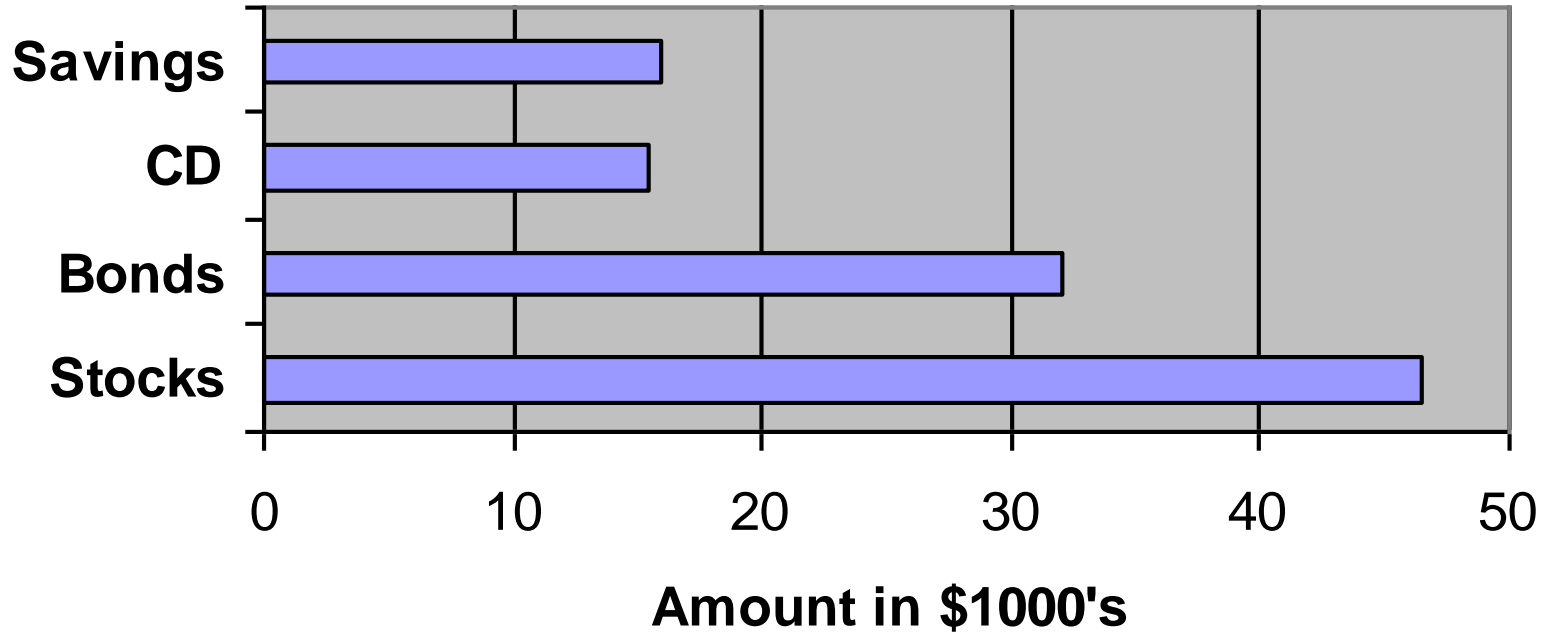
---

- Bar charts and Pie charts are often used for qualitative (category) data
- Height of bar or size of pie slice shows the frequency or percentage for each category



# Bar Chart Example 1

## Investor's Portfolio



(Note that bar charts can also be displayed with vertical bars)

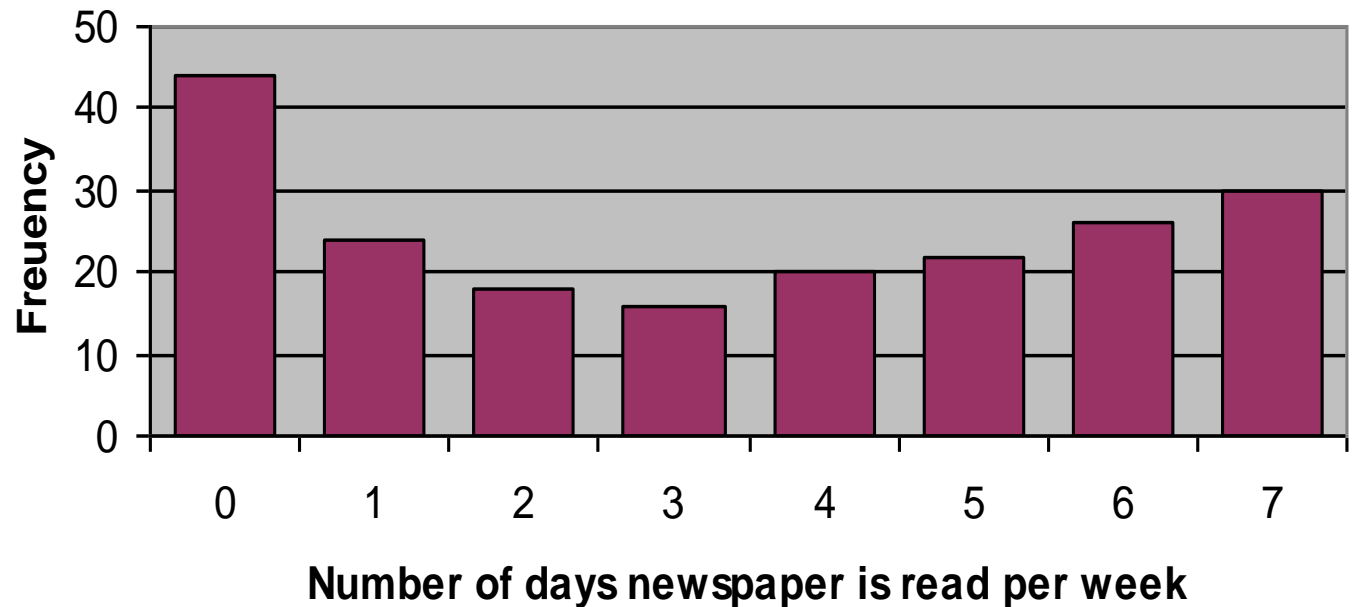




# Bar Chart Example 2

Number of days read	Frequency
0	44
1	24
2	18
3	16
4	20
5	22
6	26
7	30
<b>Total</b>	<b>200</b>

Newspaper readership per week



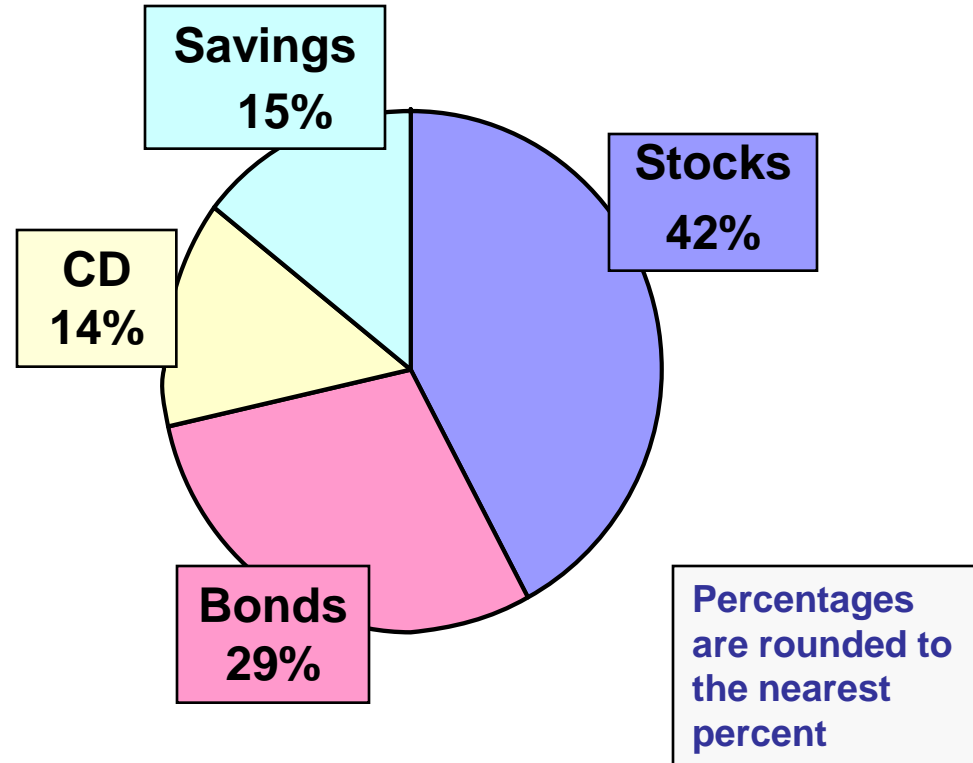


# Pie Chart Example

## Current Investment Portfolio

Investment Type	Amount (in thousands \$)	Percentage
Stocks	46.5	42.27
Bonds	32.0	29.09
CD	15.5	14.09
Savings	16.0	14.55
<b>Total</b>	<b>110</b>	<b>100</b>

(Variables are Qualitative)





# Tabulating and Graphing Multivariate Categorical Data

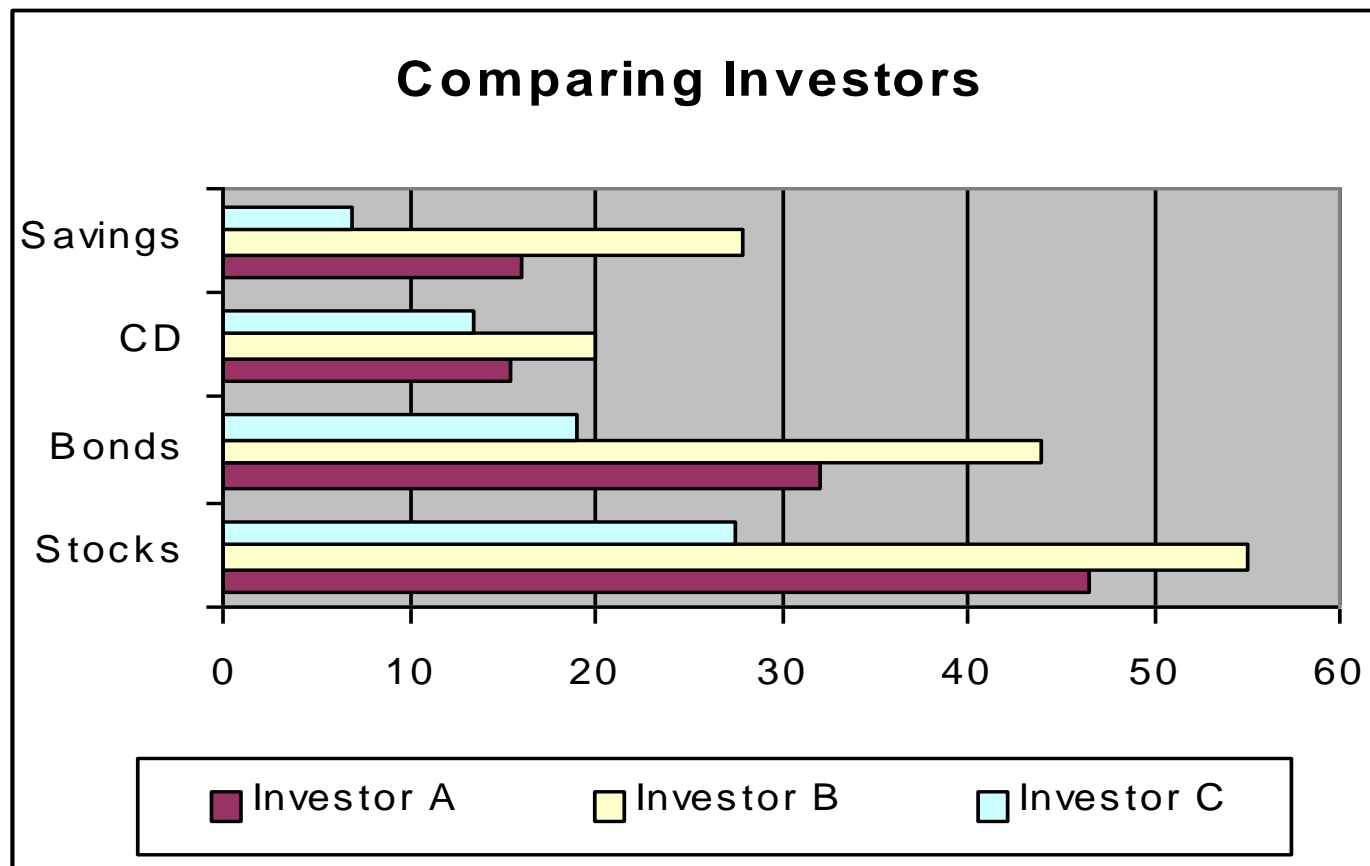
- Investment in thousands of dollars

Investment Category	Investor A	Investor B	Investor C	Total
Stocks	46.5	55	27.5	<b>129</b>
Bonds	32.0	44	19.0	<b>95</b>
CD	15.5	20	13.5	<b>49</b>
Savings	16.0	28	7.0	<b>51</b>
<b>Total</b>	<b>110.0</b>	<b>147</b>	<b>67.0</b>	<b>324</b>

# Tabulating and Graphing Multivariate Categorical Data

(continued)

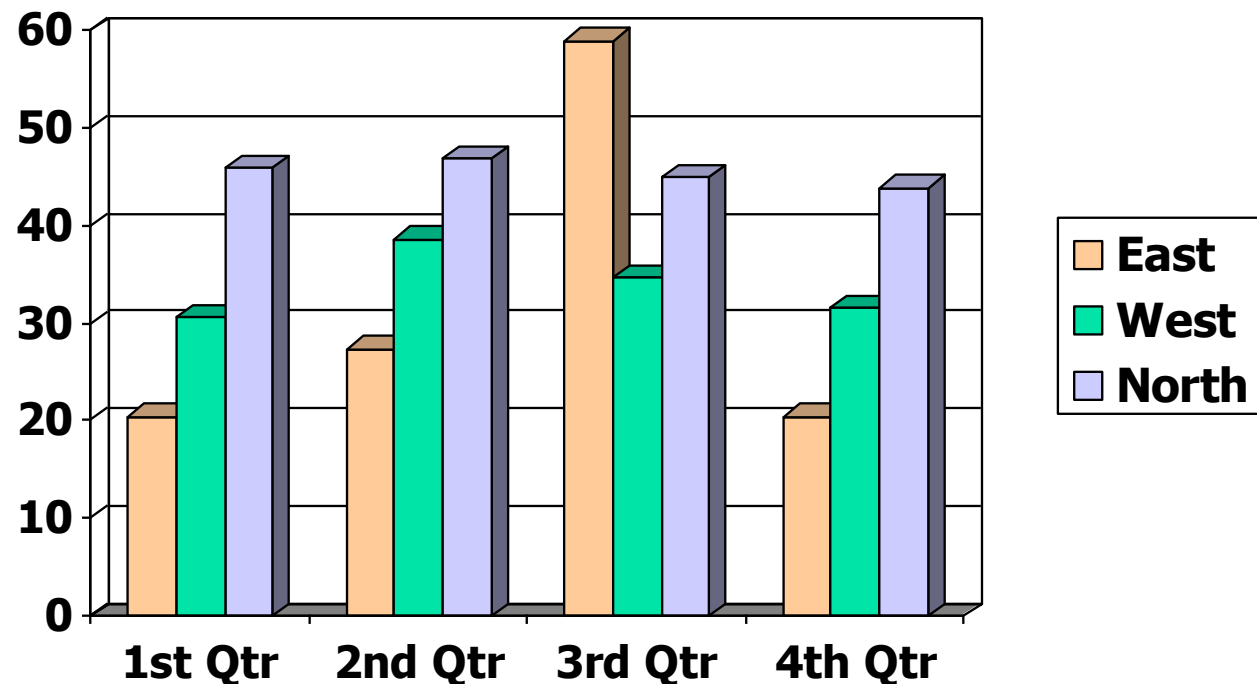
- Side by side charts



# Side-by-Side Chart Example

- Sales by quarter for three sales territories:

	1st Qtr	2nd Qtr	3rd Qtr	4th Qtr
East	20.4	27.4	59	20.4
West	30.6	38.6	34.6	31.6
North	45.9	46.9	45	43.9





# Stem and Leaf Diagram

---

- A simple way to see distribution details from qualitative data

## METHOD

1. Separate the sorted data series into leading digits (the **stem**) and the trailing digits (the **leaves**)
2. List all stems in a column from low to high
3. For each stem, list all associated leaves

# Example:

## Data sorted from low to high:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

- Here, use the 10's digit for the stem unit:

- 12 is shown as

- 35 is shown as

Stem	Leaf
1	2
3	5



# Example:

## Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 28, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

- Completed Stem-and-leaf diagram:

Stem	Leaves
1	2 3 7
2	1 4 4 6 7 8
3	0 2 5 7 8
4	1 3 4 6
5	3 8





# Using other stem units

- Using the 100's digit as the stem:
  - Round off the 10's digit to form the leaves

- 613 would become →
- 776 would become →
- . . .
- 1224 becomes →

Stem	Leaf
6	1
7	8
12	2



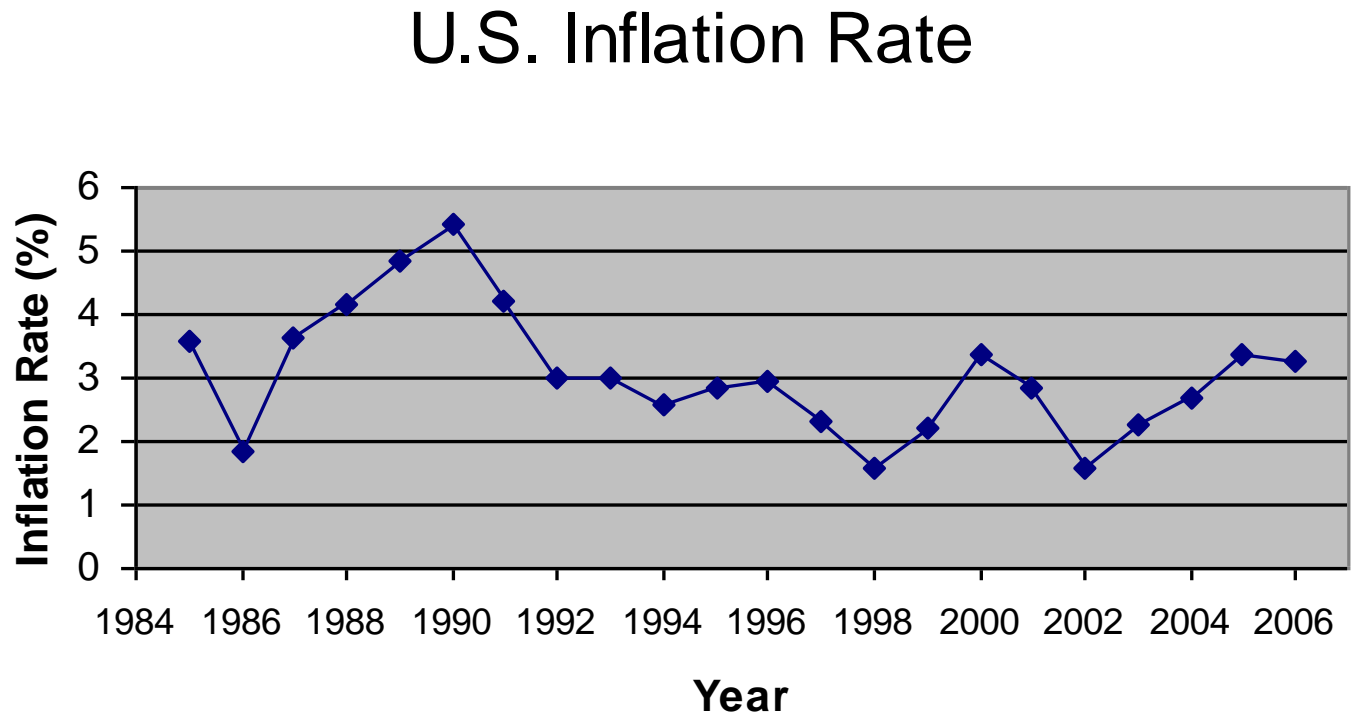
# Line Charts and Scatter Diagrams

---

- **Line charts** show values of one variable vs. time
  - Time is traditionally shown on the horizontal axis
- **Scatter Diagrams** show points for bivariate data
  - one variable is measured on the vertical axis and the other variable is measured on the horizontal axis

# Line Chart Example

Year	Inflation Rate
1985	3.56
1986	1.86
1987	3.65
1988	4.14
1989	4.82
1990	5.40
1991	4.21
1992	3.01
1993	2.99
1994	2.56
1995	2.83
1996	2.95
1997	2.29
1998	1.56
1999	2.21
2000	3.36
2001	2.85
2002	1.59
2003	2.27
2004	2.68
2005	3.39
2006	3.24

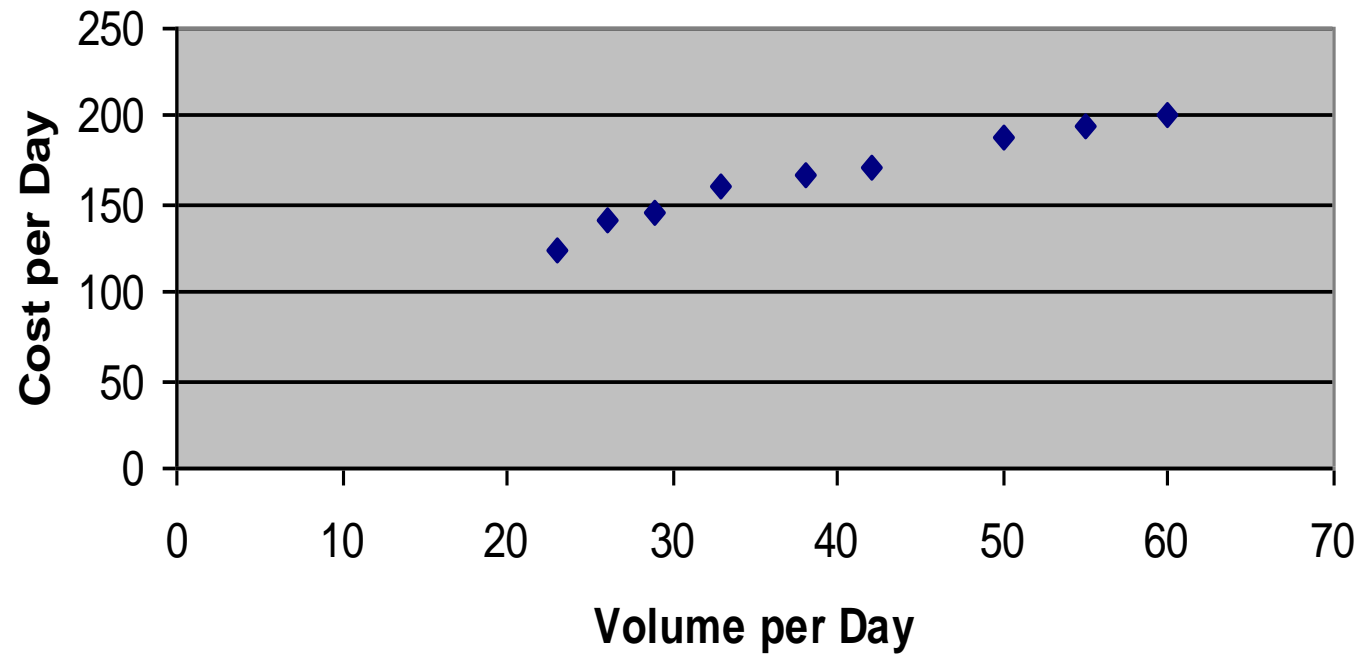




# Scatter Diagram Example

Volume per day	Cost per day
23	125
26	140
29	146
33	160
38	167
42	170
50	188
55	195
60	200

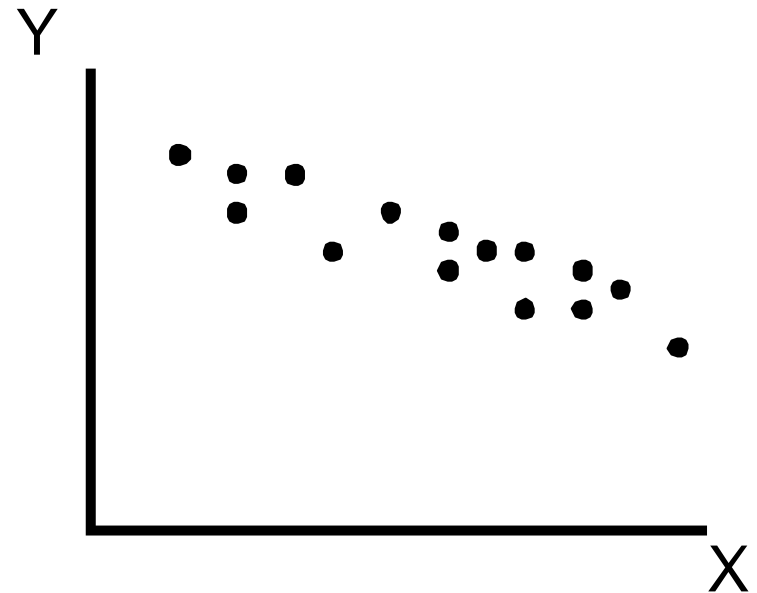
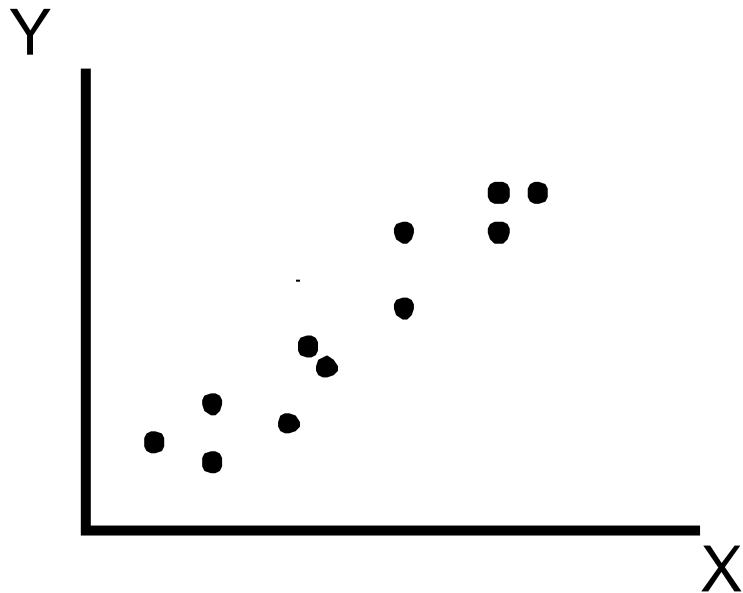
**Production Volume vs. Cost per Day**





# Types of Relationships

- Linear Relationships





# Types of Relationships

*(continued)*

- Curvilinear Relationships

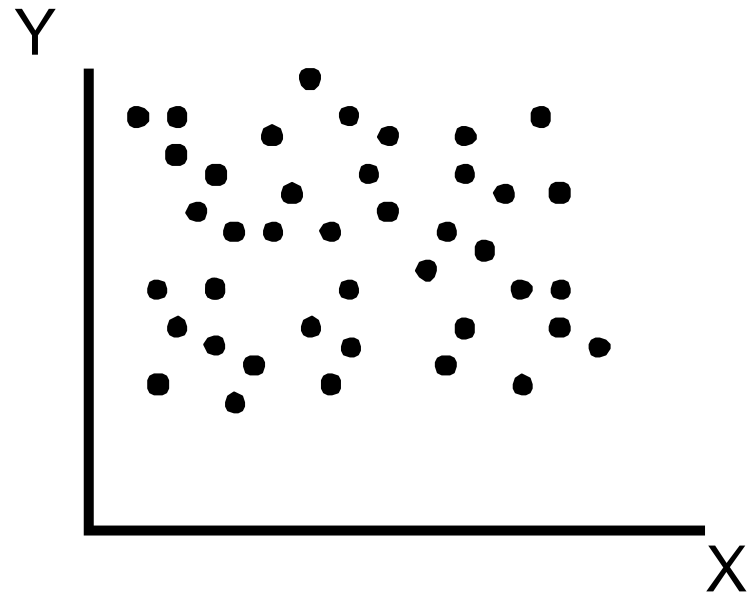
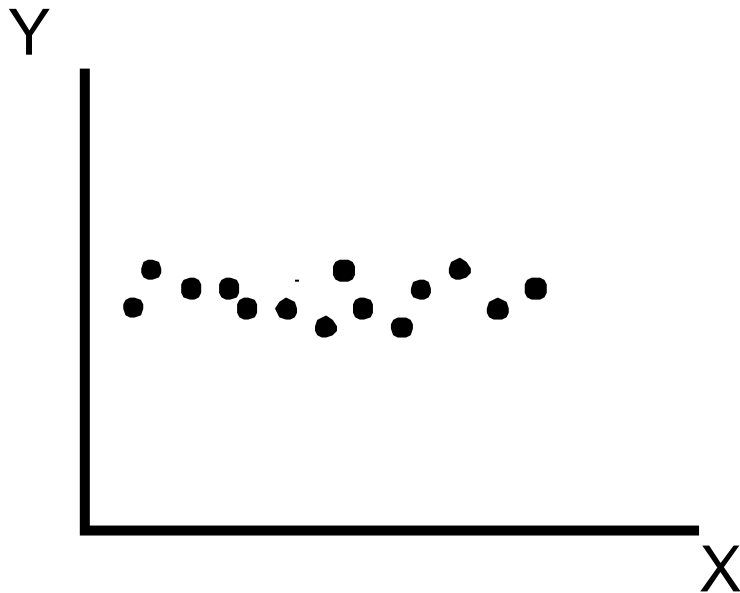




# Types of Relationships

*(continued)*

- No Relationship





# Chapter Summary

---

- Data in raw form are usually not easy to use for decision making -- Some type of organization is needed:
  - ◆ Table
  - ◆ Graph
- Techniques reviewed in this chapter:
  - Frequency Distributions, Histograms, and Ogives
  - Bar Charts and Pie Charts
  - Stem and Leaf Diagrams
  - Line Charts and Scatter Diagrams